July 5th 2012 – A11 (Red building)

Workshop on "Data Management"

Ranking and queries: as good as it gets

Davide Martinenghi (Dipartimento di Elettronica e Informazione, Politecnico di Milano - Italy)

Abstract

In today's database technology, ranking is a central aspect that captures several common needs and criteria, such as preferences, opinions, relevance, proximity, diversity, and others.

During the last years, ranking queries, i.e., queries that focus on the best results, have emerged as a paradigm of primary importance for addressing and complying with such needs and criteria.

When the query integrates several rankings into a single consensus ranking, the challenge is to obtain the top ranked results without performing a complete scan of the rankings, but rather using an "early-out" strategy.

This is particularly relevant when accessing the individual rankings is costly, as in the case of data sources on the Web.

In this talk, after illustrating the main notions and presenting the classical rank aggregation problem, I shall discuss how the field has evolved into a full-fledged area that covers most aspects of Web search and is contaminated by other fields, such as probabilistic databases and computational geometry.

Short biography

Davide Martinenghi is an Assistant Professor in the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. He received a Ph.D. in Computer Science from Roskilde University, Denmark, in 2005, with a dissertation on integrity checking in deductive databases.

His main research interests are ranking queries, data integrity maintenance, Web data management, and, in a broad sense, applications of logic to data management. He is currently also devoting his attention to the areas of reasoning in the presence of uncertainty and crowd computing.

Efficient SPARQL Query Processing over Linked Data

Katja Hose (Max Planck Institute for Informatics in Saarbrücken)

Abstract:
With the increasing popularity and the growth of the Semantic Web, more and more information becomes available on the Web in a machine readable format (RDF). By referencing data stored in remote sources, links are created that connect semantically related data in different sources. Because of the amount of data and the complexity of a user's information needs, structured query languages were developed and SPARQL became a common standard.

The challenge is how to efficiently process SPARQL queries on data distributed over multiple RDF sources. State-of-the-art approaches for linked data can roughly be classified into three categories: lookup-based query processing, federated query processing, and data warehousing -- the latter two being similar to classic data integration systems.

In the talk, we will study each of these approaches, discuss efficient techniques, and highlight differences in their applicability.

Short biography:
Katja Hose is a post-doctoral researcher at the Max Planck Institute for Informatics in Saarbrücken, Germany. She obtained a diploma (M.Sc.) in Computer Science from Ilmenau University of Technology, joined the Databases & Information Systems Group at Ilmenau University of Technology as a research associate, and received her doctoral degree in Computer Science in 2009. Afterwards, she joined the Max Planck Institute for Informatics in Saarbrücken. Her current research interests range from query processing and optimization in distributed systems and rank-aware query operators to Linked Data processing, information retrieval, and knowledge extraction.

Ontology-Based Data Access - From the Foundations to its Practical Deployment

Diego Calvanese (KRDB Research Centre for Knowledge and Data, Free University of Bozen - Italy)

Abstract:

Ontologies allow one to describe the domain of interest of an information system at a high level of abstraction. They provide mechanisms both for representing constraints holding in the underlying data and for performing automated inference over data and constraints in support of data management tasks. Ontology-based Data Access (OBDA) is a novel paradigm concerned with providing access to one or more data sources through a mediating ontology, which has gained increased attention in recent years both from the knowledge representation and from the database communities. In this talk we discuss the opportunities and challenges provided by the use of lightweight ontology languages in OBDA. We first address some of the foundational problems in this context, such as the trade-off between the expressive power of the ontology language and the efficiency of computing query answers, the impedance mismatch between the abstract objects at the ontology level and the concrete data values stored in data sources, and how to process queries over the ontology by accessing the underlying data sources. We then take into account the requirements coming from the practical deployment of ontology-based technology for data management and discuss recent results and ongoing work in thisdirection.

Short biography:

Diego Calvanese is an associate professor at the KRDB Research Centre for Knowledge and Data, Free University of Bozen-Bolzano, where he teaches graduate and undergraduate courses on knowledge bases and databases, ontologies, theory of computing, and formal languages. His research interests include formalisms for knowledge representation and reasoning, ontology languages, description logics, Semantic Web, conceptual data modeling, data integration, semi structured data management, data-aware process verification, and service modeling and synthesis. He is actively involved in several national and international research projects in the above areas, and he is the author of more than 200 refereed publications, including ones in the most prestigious international journals and conferences in Databases and Artificial Intelligence. He is one of the editors of the Description Logic Handbook. He is regularly invited to serve on the Program Committees of international conferences in the above mentioned areas and is a member of the editorial board of JAIR.

Database Algorithms and Systems for Next-Generation Hardware

Jens Thilo Teubner (Systems Group, ETH Zurich)

Abstract:

Advances in hardware technology bring computing and processing performance in ever-increasing amounts right to our fingertips. But it is increasingly difficult to harvest this performance, because most application systems have never been designed for the degrees of parallelism and hardware heterogeneity that we see today. Shifts in the hardware landscape are forcing us to re-consider basic database system designs to embrace parallelism and heterogeneity in an efficient way.

In my talk I will highlight how core database algorithms can be re-architected to match the flavors of parallelism that we see in real-world systems today. In particular, I illustrate how "handshake join" can run join operators—and thus one of the most fundamental database tasks—in a massive parallel way, while respecting the communication primitives of modern computing or networking systems. The outcome is a join implementation that exceeds the existing state-of-the-art by orders of magnitude in both performance and scalability.

Handshake join is part of a larger research agenda that I will also briefly outline in my talk. This includes novel approaches to address the heterogeneity aspect of modern hardware, e.g., by using field-programmable gate arrays (FPGAs) for database acceleration.

The work I present in this talk has been carried out in the context of the "Avalanche" research project, funded by the Swiss National Science Foundation and supported by the Enterprise Computing Center at ETH Zurich (ECC).

Short biography:

Jens Teubner is a senior researcher in the Systems Group at ETH Zurich.

His main research interest are in the use of modern computing hardware for database processing. Previously, he was a researcher at the IBM T.J. Watson Research Lab in New York, USA. He obtained his PhD at the Technische Universität München in Munich, Germany for his work on “Pathfinder”, a compiler for large-scale XQuery processing.

Taking Data Management into New Territory

Michael Grossniklaus (Portland State University)

Abstract:
The family of applications that require data management functionality is growing and evolving. Data management systems have traditionally been used for transactional data processing in automated business applications. In contrast, many emerging and novel applications focus on mobile, reactive, and analytical processing. As a consequence, the data managed by these applications and the operations used to process these data are often more complex. In this talk, we present our approach for supporting these types of applications by designing and developing general data management technologies. We illustrate our approach based on previous and ongoing projects in the area of data stream management systems. First, we discuss the design and implementation of Continuous SPARQL (C-SPARQL) query language for reasoning over streams of RDF data. Second, we describe data-driven windows, so-called frames. Unlike tuple or time-based windows, frames adapt to the data in the stream and therefore support the computation of a variety of data products. In particular, we focus on how to evaluate the quality of frame-based data products with data products computed using existing techniques. Finally, we outline how our approach can be applied to current and future domains such as mobile applications and graph data processing.

Short biography:
Michael Grossniklaus obtained his PhD in Computer Science from ETH Zurich, Switzerland in the research group of Prof. Moira C. Norrie. Currently, he is an SNSF-funded post-doc in Prof. David Maier's group at Portland State University, where he works on data stream processing and graph data management. Michael's work is situated in the area of data and information management with a focus on designing and developing database technologies to support the requirements of emerging and novel application domains. In the past, he has defined an object-oriented version model that supports context-aware data management and query processing, which has been successfully applied in a web content management system and a mobile tourist information system. As part of his previous post-doc in Prof. Stefano Ceri's group at the Politecnico di Milano, Italy, Michael has worked on stream reasoning within the "LarKC" European FP7 project and on the Panta Rhei data flow language for the Search Computing (SeCo) project.